

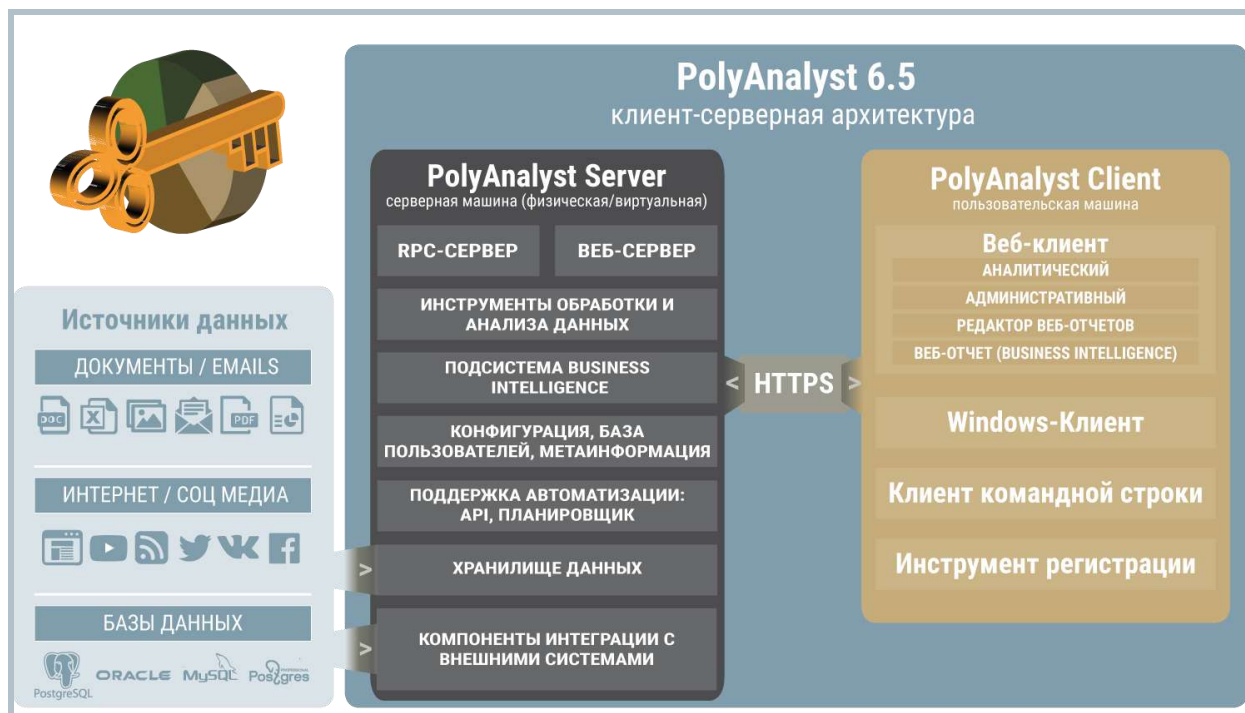
## PolyAnalyst 6.5

# ТЕХНИЧЕСКИЕ ВОЗМОЖНОСТИ И СИСТЕМНЫЕ ТРЕБОВАНИЯ



Polyanalyst - платформа визуальной разработки сценариев анализа данных и текстов. Система включает инструменты для проведения всех этапов работы с данными: загрузка из любых видов источников, подготовка и очистка, анализ с применением алгоритмов машинного обучения, а также визуализации результатов анализа и публикации интерактивных отчетов. Объединяя в одном решении механизмы анализа структурированных и текстовых данных, платформа обеспечивает исчерпывающее извлечение полезных знаний из всего существующего в организации массива информации.

# ФУНКЦИОНАЛ



## Архитектура PolyAnalyst

### Клиент-серверная архитектура

Платформа PolyAnalyst 6.5 выстроена на основе клиент-серверной архитектуры, что обеспечивает работу системы, как программного обеспечения корпоративного (enterprise) уровня. Это позволяет проведение совместной работы множества пользователей над одними аналитическими проектами и совместное использование различных связанных ресурсов, таких как сценарии анализа, словари, таксономии и многомерные матрицы. Также клиент-серверная архитектура обеспечивает высокую производительность системы за счет выполнения расчетов на наиболее мощных вычислительных ресурсах, сокращает нагрузку передаваемого трафика на корпоративную сеть, позволяет осуществлять отложенное на заданное время выполнение задач и создавать пользовательские отчеты с возможностью оповещения о наступлении событий (инцидентах) для разных групп бизнес-пользователей. Доступно централизованное управление и контроль списка пользователей, их прав и действий в системе. Клиентский доступ осуществляется через запуск приложения в ОС, или через "тонкий клиент" в веб-браузере.

### Безопасность

Принимая во внимание, что сегодня одним из наиболее ценных и чувствительных активов современной организации являются данные, PolyAnalyst обладает надежными механизмами для обеспечения безопасности данных. Связь между пользовательским клиентом и сервером осуществляется в зашифрованном виде с ключом шифрования, заново генерируемым сервером для каждого сеанса связи. Защищенная авторизация в PolyAnalyst осуществляется в соответствии с доступными правами пользователя и с применением многосложного пароля. Система администрирует как индивидуальные, так и групповые пользовательские права, и контролирует всю последовательность действий, проводимых пользователем в программе.

## Масштабируемость и производительность

PolyAnalyst обеспечивает масштабируемость промышленного уровня: система способна обрабатывать индустриальные объемы данных в кратчайшие временные интервалы. Такая масштабируемость обеспечивается за счет комбинации нескольких факторов. PolyAnalyst использует жесткие диски вместо запоминающих устройств с произвольным доступом (RAM) для хранения всех типов данных, а также поддерживает применение метаданных. В системе реализуются специальные масштабируемые аналитические алгоритмы, что позволяет обрабатывать массивные объемы данных. Пользовательский интерфейс PolyAnalyst позволяет аналитику разрабатывать сложные сценарии анализа без загрузки данных в систему, тем самым, экономя время пользователя. Исполнение PolyAnalyst в качестве 64-разрядного приложения и возможность ее имплементации в вычислительную систему из объединенных в группы серверов (server farm) демонстрирует высочайшую производительность системы и позволяет сотням пользователей одновременно работать с данными и просматривать результирующие отчеты.

## Повторяемые сценарии анализа

Типовой аналитический сценарий включает последовательность этапов загрузки, подготовки, анализа данных и конструирования отчетов. Зачастую, такая последовательность аналитических шагов выполняется с определенной регулярностью на вновь загружаемых массивах данных. PolyAnalyst - это самодокументируемая (self-documenting) система, включающая интуитивно понятные визуальные инструменты разработки и редактирования многошаговых, повторяемых сценариев анализа данных. Работа с платформой не требует профессиональных знаний программиста или аналитика. Учиться ее применению легко и увлекательно. Проведение анализа и выгрузка отчета могут быть запланированы для повторного выполнения в определенный пользователем момент времени. Это гарантирует, что бизнес-пользователи будут иметь своевременный доступ к аналитическим отчетам, основанным на наиболее актуальных данных.

## Загрузка и интеграция данных

Платформа PolyAnalyst способна загружать и интегрировать данные практически из любых существующих источников. Система может работать с разрозненными источниками, включая все популярные базы данных, статистические источники и электронные таблицы. Кроме того, PolyAnalyst работает с документами нескольких десятков форматов таких, как html, doc, pdf и txt, и собирает данные из интернет-ресурсов, веб-сайтов, RSS-каналов и социальных сетей. PolyAnalyst включает визуальный интерфейс интеграции и объединения данных из разрозненных источников для дальнейшего создания витрин данных и проведения анализа. Также возможно догружать данные или извлекать их из ранее созданных проектов в PolyAnalyst.

## Предобработка данных

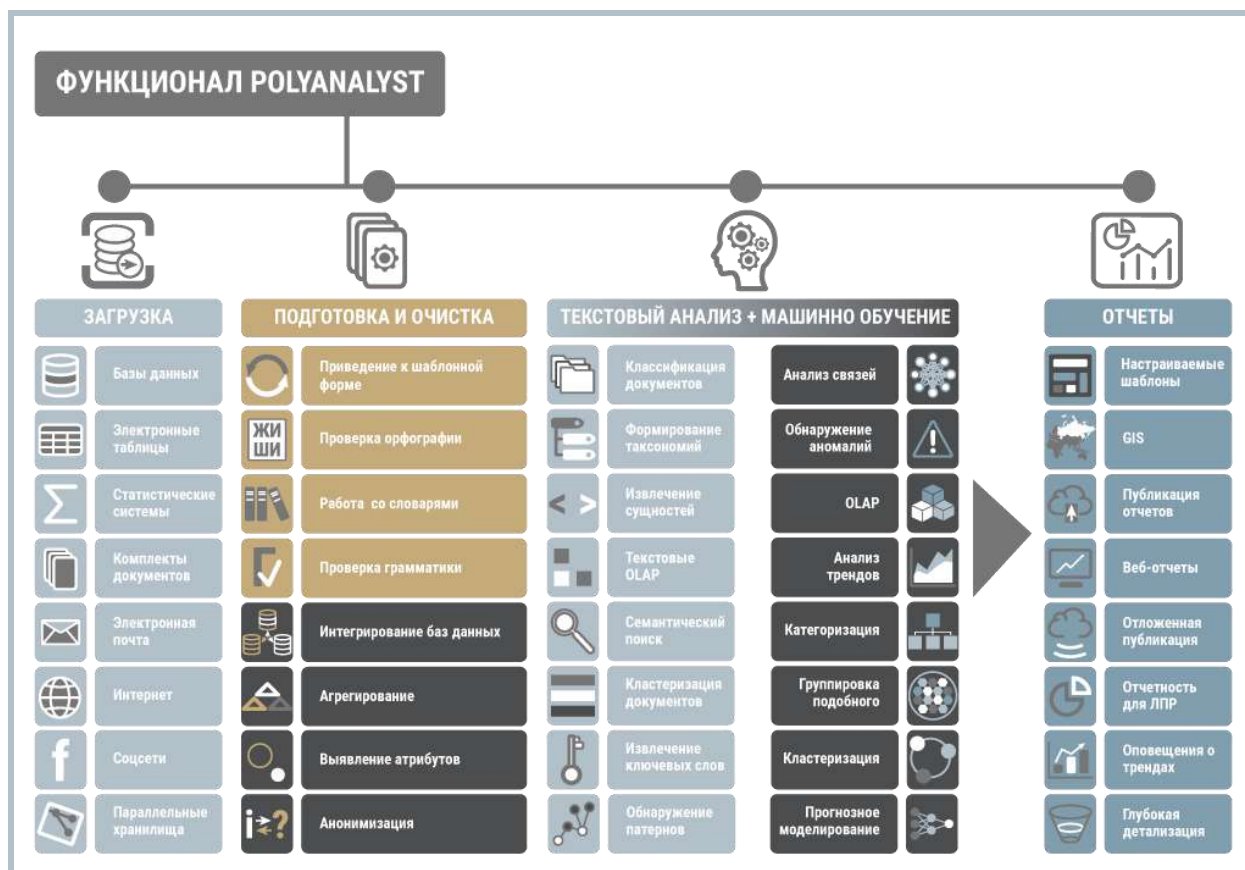
Каждый проект анализа данных начинается с их нормализации, предобработки и разведочного анализа. Подавляющая часть времени аналитика тратится на предобработку данных, предшествующую применению алгоритмов машинного обучения. PolyAnalyst предоставляет аналитику широкий набор мощных инструментов предобработки, проводящих очистку, агрегацию и вывод новых атрибутов. Практически любая задача по преобразованию данных может быть решена за счет применения инструментов последовательной предобработки PolyAnalyst.

## Алгоритмы анализа

Важнейшим значением для пользователя, работающего с системами обнаружения и извлечения знаний, является возможность обучения математических алгоритмов на исторических данных и высокоточного прогнозирования результатов будущих событий. PolyAnalyst предлагает широкий набор аналитических алгоритмов для кластеризации и категоризации данных, прогнозирования, анализа связей, определения паттернов и обнаружения аномалий в данных. Пользователь способен легко решать аналитические задачи, применяя множество доступных аналитических алгоритмов PolyAnalyst таких, как нейронные сети и деревья решений, Байесовские классификации и метод опорных векторов, CHAID-анализ и логистическая регрессия, метод рассуждения на основе прецедентов и эволюционное программирование.

## Текстовый анализ

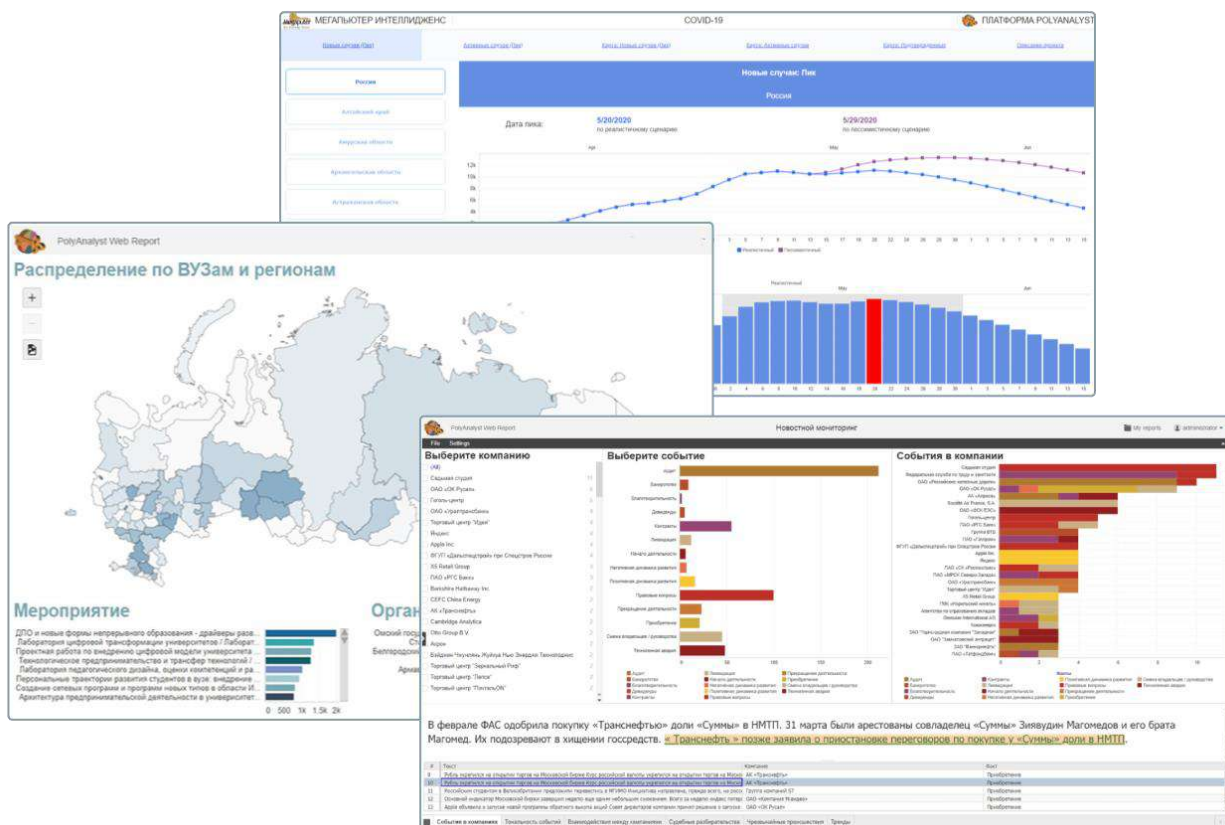
Данные могут быть представлены не только в структурированной, но и неструктурированной текстовой форме. Они расположены во всем множестве документов, находящихся в информационном обороте организации: счета, акты, договоры, анкеты, сводки, обращения, НПА и множество прочих. Также огромный пласт полезной информации содержится в текстовых данных, расположенных в Интернете. PolyAnalyst предлагает объемную коллекцию алгоритмов анализа текстов, позволяющую пользователю решить все множество возможных аналитических задач. Основанный на базе собственной лингвистической платформы и объединяющий множество семантических словарей PolyAnalyst представляет собой передовой инструмент обработки естественного языка (Natural language processing, NLP). С помощью PolyAnalyst возможно проводить как обуславливаемый пользователем анализ, к примеру категоризация на основе таксономий или текстовый OLAP, так и анализ, обуславливаемый данными, например интеллектуальную проверку орфографии, извлечение ключевых слов и сущностей, кластеризации и построение таксономий.



### Ключевые возможности PolyAnalyst

## Многомерный анализ

Разработка многомерных кубов на основе проанализированных данных позволяет пользователю ответить на различные бизнес-вопросы, распределяя данные в многомерной системе координат. Данная технология широко известна как OLAP (Оперативный анализ данных). PolyAnalyst предлагает надежный механизм OLAP в качестве одного из алгоритмов анализа данных. Более того, PolyAnalyst развивает стандарт OLAP, добавляя измерения, определяемые полями, содержащими неструктурированный текст, и предлагая уникальный многомерный интерфейс OLAP.



## Веб-отчеты PolyAnalyst

### Интерактивные визуализации и отчеты

PolyAnalyst предоставляет пользователям средства отображения результатов сложных алгоритмов анализа данных в виде интерактивных графических отчетов, доступных для понимания неспециалистам. Они, как правило, представляют из себя детализированную смесь графиков, таблиц, чисел, текста и ссылок на прочие объекты PolyAnalyst. Отчеты могут быть запланированы на обновление в заданное время для предоставления бизнес-пользователям наиболее актуальных результатов анализа. Конструктор визуализаций обладает интуитивным пользовательским интерфейсом, поддерживающим функцию drag-and-drop. Имеются широкие возможности по выбору и настройке цветовой схемы, типа визуализации, области детализации, отображения логотипа компании и т. п. Созданные отчеты доступны в любом веб-браузере. Имеется возможность скачивать отчет в удобном электронном формате, открывать к нему доступ другим сотрудникам, распечатывать его для руководства, коллег или клиентов.



# ТЕХНИЧЕСКОЕ ОПИСАНИЕ

## 1. Загрузка данных

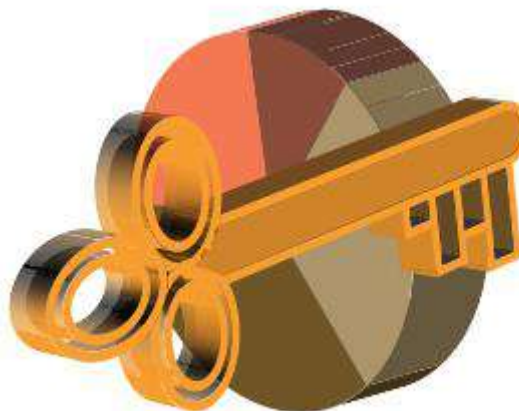
- a. Источники данных: ODBC, OLEDB, XML, CSV, MS Excel, Web, Файловые системы, FTP
- b. Документы различных форматов: PDF, ASCII, HTML, MS Word, MS RTF, RSS Feeds
- c. Форматы символов: ASCII, Latin-1, Double-byte, UTF
- d. Веб-ресурсы
- e. Прямой импорт данных электронной почты (PST or NSF)
- f. Импорт данных с сервера электронной почты
- g. Соцмедиа:
  - i. Twitter и Facebook API
  - ii. Webhose, BrandWatch и SM2 API
- h. Ссылки на дата-сети
- i. Федеративный поиск

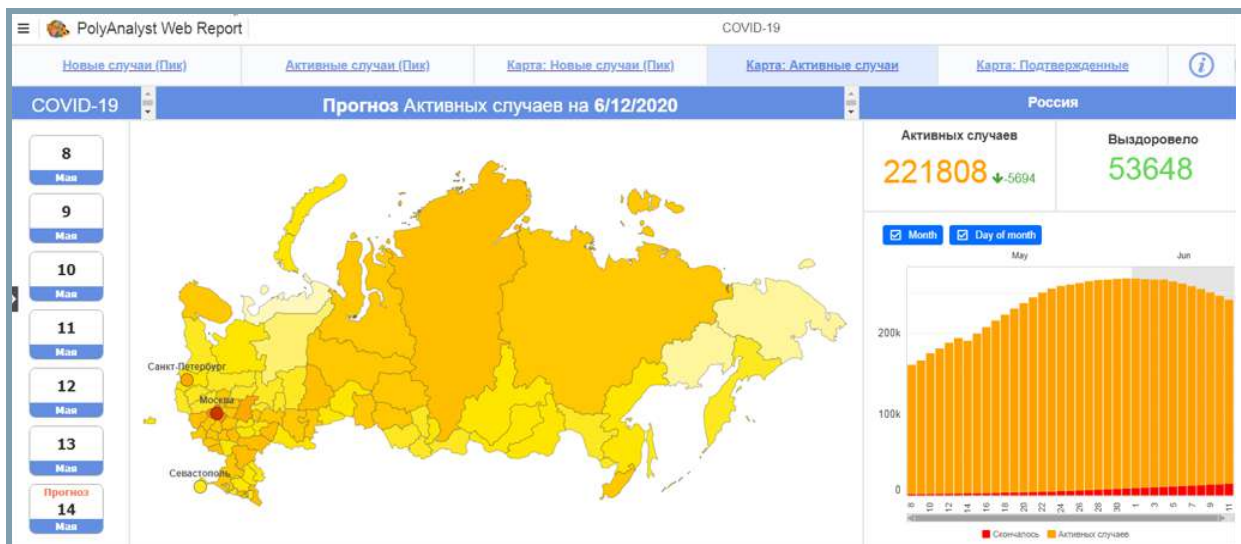
## 2. Интеграция данных

- a. Присоединение данных по наборам ключей
- b. Объединение данных
- c. Извлечение данных из других проектов PolyAnalyst
- d. Экспорт результатов анализа во внешние реляционные СУБД

## 3. Очистка, предобработка данных и разведочный анализ

- a. Сопоставление имени, типа и значения атрибута
- b. Агрегирование данных
- c. Консолидирование данных
- d. Преобразование данных
- e. Выборка и разделение
- f. Вывод новых атрибутов
- g. OLAP
- h. Статистика
- i. Анализ связей
- j. Геопространственное распределение данных





### Геопространственное распределение данных в PolyAnalyst – уровень регионов РФ

#### 4. Прогнозное моделирование и машинное обучение

##### a. Классификация

- i. Деревья решений
- ii. Наивный байесовский классификатор
- iii. Бустинг
- iv. Байесовская сеть
- v. Random forest
- vi. CHAID
- vii. Рассуждения на основе прецедентов
- viii. Нейронные сети
- ix. Метод опорных векторов
- x. Обучение на примерах
- xi. Логистическая регрессия
- xii. MARS-Многомерные адаптивные регрессионные сплайны

##### b. Кластеризация

- i. CHAID
- ii. Локализация аномалий
- iii. Самоорганизующаяся карта Кохонена

##### c. Прогнозирование

- i. Нейронные сети
- ii. Эволюционное программирование
- iii. Линейные регрессии

##### d. Обнаружение аномалий

- i. Анализ распределений
- ii. Аудит данных

##### e. Группирование

- i. Анализ рыночной корзины
- ii. Визуальный анализ связей
- iii. Дискриминантный анализ рыночной корзины

##### f. Снижение размерности

- i. Факторный анализ
- ii. Метод главных компонент
- iii. Связывание записей (Fuzzy Matching)
- iv. Анализ корреляций

##### g. Анализ временных рядов

- i. Аддитивный анализ временных рядов

##### h. Сетевой анализ

- i. Анализ соцсетей
- ii. Анализ связей

## 5. Интеллектуальный анализ текстов

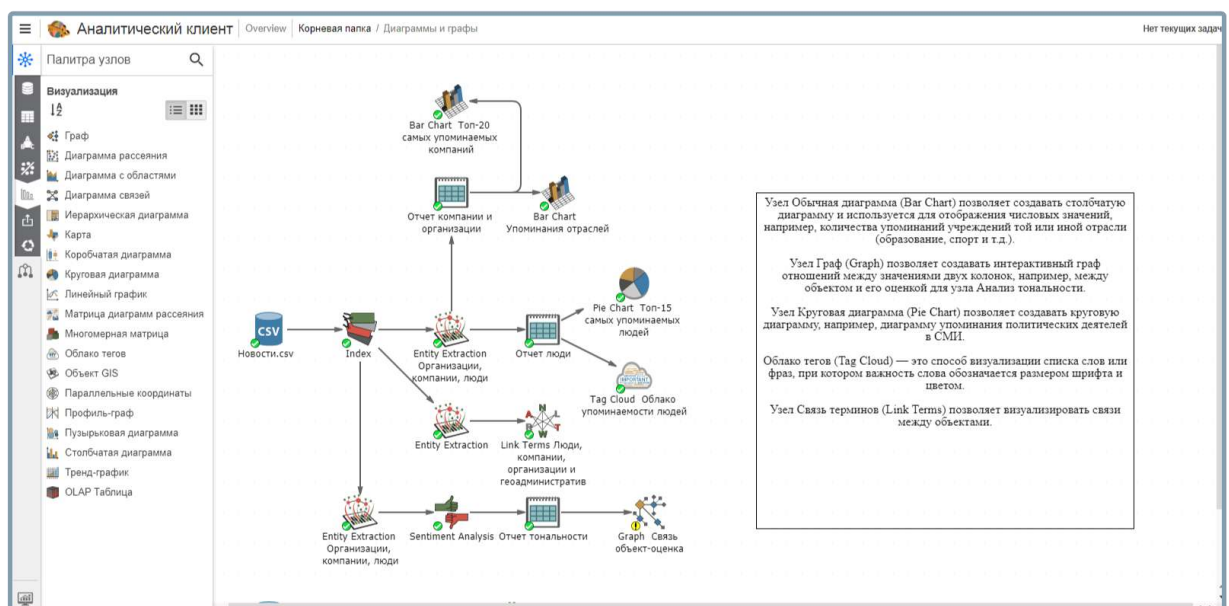
- a.** Очистка и нормализация текстовых данных
    - i.** Обнаружение дубликатов
    - ii.** Фрагментный анализ
    - iii.** Редактируемые словари фраз, стоп-слов, синонимов и прочих семантических отношений
    - iv.** Интеллектуальная проверка и коррекция орфографии
  - b.** Пользовательский анализ
    - i.** Обнаружение информации с помощью специализированных поисковых запросов
    - ii.** Расширение запроса с помощью семантических взаимосвязей (синонимы, гипонимы и т. д.)
    - iii.** Категоризация на основе таксономий
    - iv.** Автоматизированная категоризация на основе заданных примеров
    - v.** Извлечение стандартных сущностей
    - vi.** Извлечение специализированных сущностей
    - vii.** Извлечение отношений
    - viii.** Поиск схожих документов
    - ix.** Многомерные отчеты о результатах текстового анализа (Text OLAP)
  - c.** Анализ, обусловленный данными
    - i.** Определение языка
    - ii.** Определение части речи
    - iii.** Извлечение ключевых слов
    - iv.** Извлечение фактов
  - v.** Синтаксический анализ: анализатор зависимостей
  - vi.** Синтаксический анализ: избирательный анализатор
  - vii.** Извлечение стандартных сущностей
  - viii.** Текстовая кластеризация (бинарная и иерархическая)
  - ix.** Формирование таксономий из анализа необработанных данных
  - x.** Сентиментный анализ
  - xi.** Визуальный кластерного анализа
  - xii.** Модель классификации, обученная на основе предварительно классифицированных примеров
- d.** Языки
    - i.** Поддерживается обработка текстов на 16 языках: Английский, Испанский, Французский, Португальский, Русский, Немецкий, Итальянский, Голландский, Польский, Греческий, Турецкий, Арабский, Китайский, Японский, Корейский, Тайский
  - e.** Уникальный функционал
    - i.** Обнаружение отрицаний
    - ii.** Разрешение анафор
    - iii.** Обнаружение модальности
    - iv.** Расшифровка аббревиатур
    - v.** Язык определения паттернов (PDL)
    - vi.** Регулярные выражения



- j. Кастомизация: создание и поддержка специализированных отраслевых словарей
- k. Кастомизация: создание и поддержка специализированных отраслевых таксономий
  - i. Финансы
  - ii. Страхование
  - iii. Медицина
  - iv. Производство
  - v. ТЭК
  - vi. Телеком
  - vii. HoReCa

**6. Графические интерактивные визуализации результатов анализа с функцией глубокой детализации данных**

- a. Гистограммы (2D и 3D)
- b. Точечные диаграммы (рассеяния)
- c. Линейные графики
- d. Пузырьковые диаграммы
- e. Тепловые карты
- f. Графики правил
- g. Диаграмма связей
- h. Облака тегов
- i. Иерархические диаграммы
- j. Многомерные матрицы
- k. Графики трендов



**Интерфейс PolyAnalyst и пример повторяемого сценария анализа данных**

**7. Формирование отчетов**

- a. Настраиваемые шаблоны отчетов
- b. Обновление и актуализация данных в отчетах
- c. Интерактивность и глубокая детализация данных
- d. Доступ к отчетам через веб-браузеры

**8. Планирование**

- a. Автоматический запуск выполнения сценариев анализа в заданное время
- b. Оповещения при наступлении заданного условия или при инциденте

## **9. Аналитические сценарии**

- a.** Визуальная разработка повторяемых сценариев анализа
- b.** Drag-and-drop интерфейс конструктора отчетов
- c.** Функция группирования узлов для управления массивными
- d.** Сценарии

## **10. Масштабируемость**

- a.** Клиент-серверная архитектура
- b.** Кластерная архитектура для большим объемов данных
- c.** Клиент-серверное взаимодействие через http/https протоколы
- d.** Использование жесткого диска вместо оперативной памяти
- e.** Масштабируемая реализация алгоритмов анализа
- f.** Исполнение в качестве 64-разрядного приложения
- g.** Разработка сценария анализа без загрузки данных

## **11. Простота использования**

- a.** Интерактивный drag-and-drop интерфейс всех модулей платформы
- b.** Визуальная разработка повторяемых сценариев анализа
- c.** Простая интеграция результатов аналитики в отчетность
- d.** Визуальное создание красивых интерактивных отчетов для бизнес-пользователей
- f.** Публикация отчетов в популярные форматы электронных документов
- g.** Группирование узлов сценария
- h.** Разделение узлов на предметные области

## **12. Безопасность**

- a.** Аутентификация на основе имени пользователя и многосложного пароля
- b.** Поддержка LDAP и MS Windows аутентификации
- c.** Полностью зашифрованное клиент-серверное взаимодействие
- d.** Регистрация всех действий пользователя

# МИНИМАЛЬНЫЕ ТРЕБОВАНИЯ



## PolyAnalyst Server

**Для сервера PolyAnalyst 6.5** требуется 64-битная версия Microsoft Windows 2008R2+SP1(6.1.7601)/7+SP2/8/2012(6.2.9200)/10

**Для веб-отчетов PolyAnalyst** требуются следующие веб-браузеры, начиная с указанных версий:

- Microsoft Edge
- GoogleChrome 59
- Mozilla Firefox 48
- Safari 10

**Cookies и JavaScript** должны быть включены в вашем браузере.

**Для работы с веб-отчетами** вы можете использовать и другие браузеры, однако функциональность веб-отчетов может быть неполной.

**Для работы 64-битного сервера PolyAnalyst 6.5** необходимо 4096 МБ оперативной памяти и 4GB свободного места на жёстком диске.

## PolyAnalyst Client

**Для установки клиентских приложений PolyAnalyst** требуется 64-битная версия Windows 2008R2+SP1(6.1.7601)/7+SP2/8/2012(6.2.9200)/10

**Для веб-отчетов и доступа через тонкий клиент PolyAnalyst** требуются следующие веб-браузеры, начиная с указанных версий:

- Microsoft Edge
- GoogleChrome 59
- Mozilla Firefox 48
- Safari 10
- Яндекс.Браузер

**Cookies и JavaScript** должны быть включены в вашем браузере.

**Для работы с веб-отчетами** вы можете использовать и другие браузеры, однако функциональность веб-отчетов может быть неполной.

**Для работы клиентских приложений PolyAnalyst** необходимо 3072 МБ оперативной памяти (2048 для 32-битной операционной системы) и 1GB свободного места на жёстком диске.